Genome-Wide Association Studies of Allergic Diseases

Mayumi Tamari¹, Shota Tanaka¹ and Tomomitsu Hirota¹

ABSTRACT

Allergic diseases are complex diseases caused by a combination of genetic and environmental factors. To determine the genetic components of these diseases and to discover the genes and cellular pathways underlying them, a large number of genetic studies have been conducted. Progress in genetics enables us to conduct genome-wide association studies (GWASs), which is a comprehensive and unbiased approach to identify susceptibility loci for multifactorial diseases. Recent GWASs have convincingly detected a large number of loci associated with allergic diseases. Candidate genes in the susceptibility loci suggest roles for epithelial barrier functions, innate-adaptive immunity, IL-1 family signaling, regulatory T cells and the vitamin D pathway in the pathogenesis of allergic diseases. Interestingly, the IL1RL1, HLA, IL13 and C11orf30 regions are overlapping susceptibility loci among atopic dermatitis and asthma or allergic rhinitis. Although a more complete collection of associated genes and pathways is needed, biologic insights revealed by GWASs improve our understanding of the pathophysiology of human allergic diseases and contribute to the development of better treatment and preventive strategies.

KEY WORDS

asthma, atopic dermatitis, genetics, Genome-Wide Association Study, polymorphisms

INTRODUCTION

The genome is the set of all genes, regulatory sequences and other information included in the non-coding regions.1,2 The Human Genome Project is a coordinated international project with the primary goal of determining the consensus sequence of the human genome.1,2 After the draft sequence of the human genome was reported in 2001, researchers focused on the genomic variation among individuals.3-6 A base variation is referred to as a single-nucleotide polymorphism (SNP), and humans have a heterozygous site roughly every 300 bases in their genome.5 It has been suggested that there are about 10 million SNPs across the genome and one million SNPs in gene regions.5 Some of them, especially around genes, are considered to influence variations in the timing, amount, or function of a protein produced from a gene and to contribute to disease risk. Most common diseases are caused by the interaction of genetic and environmental factors, and genetic variations are a component of the genetic factors. To discover disease-related gene variations, a large number of case-control studies, which compare the frequencies of alleles or genotypes between cases and controls, have been conducted.7 Since the 1990s, association studies for allergic diseases using the candidate gene approach have been performed in various ethnic populations.8 The candidate gene approach is hypothesis-driven and the findings are easy to interpret; however, these studies cannot identify novel susceptible genes or pathways. Furthermore, some of the studies had issues such as the lack of statistical power and replication study, insufficient polymorphism coverage of a gene, and small sample sizes.9 The ‘common disease-common variant’ hypothesis predicted that common genetic variants could play a role in the etiologies of common diseases.10 The genome-wide association study (GWAS) is a test of the association between common genetic variants spreading across the genome and disease.11,12 Comprehensive, well-powered, genome-wide surveys us-
ing GWAS have revealed disease susceptibility loci.11,12 GWASs have accelerated biomedical research to discover the genes and cellular pathways underlying disease in an unbiased and hypothesis-free manner.7 This review focuses on GWAS of allergic diseases, and we also discuss the road ahead for the genetics of allergic diseases.

**GENOME-WIDE ASSOCIATION STUDIES**

The international HapMap project (http://hapmap.ncbi.nlm.nih.gov/) started in 2002 to develop a public database that could help researchers find genes associated with human diseases and individual responses to pharmacological agents.5,6 The HapMap project focuses only on common SNPs with minor allele frequencies >1% in the population and defines these patterns across the whole genome. The discovery of the haplotype structure of the human genome indicated that a limited set of 1,000,000 SNPs could capture ~90% of the genetic variation in the population.7 The data from the HapMap project and the development of dense genotyping arrays have enabled us to perform GWASs on a large number of samples. Genotyping arrays have been developed with reference to linkage disequilibrium, and are now able to assay up to ~2 million variants. However, if large numbers of SNPs are assessed simultaneously in a GWAS, the statistical significance threshold must be adjusted for type 1 errors (false positive). Generally, a genome-wide significance threshold level of 5 × 10^-8 (0.05/1,000,000) is applied if a test of independence is performed for 1,000,000 SNPs with a conventional P value < 0.05 for significance.11

Genotype imputation is a statistical approach deducing the allelic states at polymorphic sites not actually genotyped from the surrounding SNP data, based on linkage disequilibrium and likelihood estimates.13 Recently, a number of GWAS consortiums have been founded with the aim of conducting rigorous statistical and comprehensive GWAS meta-analyses for common disorders. Imputation is useful for merging distinct GWASs using different genotyping platforms and enables powerful combined analyses of GWASs for the discovery of novel associations.13,14 A collaborative Transnational Asthma Genetics Consortium (TAGC) bringing GWASs from all over the world together and completing a meta-analysis for asthma examining over 30,000 cases and 50,000 controls is now ongoing.

**GWASs OF ASTHMA**

Bronchial asthma is a heterogeneous disease caused by a combination of genetic and environmental factors.8 A large number of genetic studies using the candidate gene approach have been conducted to discover the genes and cellular pathways underlying asthma.8,9,15 Recently, systemic, well-powered, genome-wide surveys using GWASs have explored the relationship between SNPs and asthma susceptibility. The findings of GWASs of asthma imply the importance of genes that play a role in communication of epithelial damage to the adaptive immune system and activation of airway inflammation.8,9,15

The first GWAS of asthma identified a locus on chromosome 17q21.1 that contributed to the risk of childhood asthma.16 Genetic variants in the locus were strongly correlated in cis with transcript levels of ORMDL3 in Epstein-Barr virus (EBV)-transformed lymphoblastoid cell lines from children with asthma. A number of studies in various ethnic populations replicated the association; however, it appears that the region is associated with childhood, but not adult-onset, asthma.17,22

Sleiman et al. conducted a GWAS involving children with persistent asthma who required daily inhaled glucocorticoid therapy and matched controls, and identified a novel asthma-susceptible locus on chromosome 1q31, which contains CRB1 and DENND1B. The latter is expressed in dendritic cells and activates T cells.23

A recent large-scale, consortium-based GWAS of asthma was performed with 10,365 cases and 16,110 controls and identified six susceptible loci containing IL1RL1/IL18R1, HLA-DQ, IL33, SMAD3, ORMDL3/GSDMB and IL2RB.24 The ligand for IL1RL1 is IL-33, which has been shown to be important in Th2-associated disease models.25 IL-33 is considered to be important for host defense against nematodes by inducing T helper type 2 (Th2) cytokine production via IL-33 receptors and is also a key cytokine for inducing and activating innate lymphoid cells.26 These findings imply that genetic factors involved in the IL-33 pathway play an important role in the pathophysiology of human bronchial asthma.

In 2011, a meta-analysis of GWASs of asthma in ethnically diverse North American populations was reported.27 The study examined reports on individuals of European American, African American or African Caribbean, and Latino ancestry and identified a total of five susceptibility loci. Four loci on 17q21, near IL1RL1, TSLP and IL33 were already reported regions.24,28 A new asthma susceptibility locus at PYHIN1 was identified with the association being specific to individuals of African descent.

In 2011, we conducted a GWAS of adult asthma in the Japanese population and reported five candidate loci.29 The most significant association with adult asthma was observed at rs404860 in the major histocompatibility complex (MHC) region on chromosome 6p21, which is close to rs2070600 previously reported for association with FEV1/FVC by GWASs for lung function.30,31 Reduction of FEV1/FVC is a characteristic of obstructive lung diseases such as asthma. There might be an important common genetic determinant of lung function in both healthy individuals and asthma at the locus. An associated re-
region on chromosome 5q22 includes the TSLP gene. TSLP derived from epithelial cells induces Th2 cell-type inflammation and is involved in the pathogenesis of asthma. Interestingly, the TSLP/WDR36 locus is associated with bronchial asthma across ethnic lines with genome-wide significance. A recent study has shown that treatment with corticosteroids and salmeterol synergistically suppresses poly(I:C)-induced TSLP in airway epithelial cells. Thus, TSLP might meterol synergistically suppresses poly(I:C)-induced

Table 1 Susceptibility loci of atopic dermatitis identified by GWAS

<table>
<thead>
<tr>
<th>Populations</th>
<th>Ref.</th>
<th>Chromosome</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>European</td>
<td>ref. 46</td>
<td>11q13.5</td>
<td>C11ORF30/LRRC32 (GARP)</td>
</tr>
<tr>
<td>Chinese</td>
<td>ref. 47</td>
<td>1p21.3</td>
<td>FLG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5q11.1</td>
<td>TMEM232/SLC25A46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20q13.3</td>
<td>TNFRSF6B/ZGPAT</td>
</tr>
<tr>
<td>European</td>
<td>ref. 48</td>
<td>11q13</td>
<td>OVL1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19p13.2</td>
<td>ACTL9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5q31</td>
<td>KIF3A/IL4/IL13</td>
</tr>
<tr>
<td>Japanese</td>
<td>ref. 53</td>
<td>2q12</td>
<td>IL1RL1/IL18R1/IL18RAP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3p21.33</td>
<td>GLB1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3q13.2</td>
<td>CCDC80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6p21.3</td>
<td>the MHC region</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7p22</td>
<td>CARD11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10q21.2</td>
<td>ZNF365/EGFR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11p15.4</td>
<td>OR10A3/NLRP10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20q13</td>
<td>CYP24A1/PFDN4</td>
</tr>
</tbody>
</table>

FEV1-pre-FEV1i) in response to inhaled corticosteroids. A functional GLCCI1 variant is associated with substantial decrements in the inhaled corticosteroid response in asthma. The low responder allele shows significantly lower enhancer activity than the other allele. Interestingly, GLCCI1 is induced by corticosteroids in human B cells in response to costimulation with IL-4 and anti-IgM. The findings suggest that GLCCI1 is a target molecule of glucocorticoid in B cells under the Th2 condition.

GWASs of atopic dermatitis

Atopic dermatitis is a chronic skin disease involving skin barrier dysfunctions and cutaneous inflammatory hypersensitivity with a strong genetic basis. Patients with atopic dermatitis are particularly susceptible to a number of microbial organisms, and pruritus is a major symptom of atopic dermatitis. The itch-scratch cycle can lead to damage to the epidermal keratinocytes. Table 1 summarizes the findings from recent GWASs of atopic dermatitis, and genes within identified atopic dermatitis susceptibility loci are shown. A recent study has reported that common loss-of-function variants in FLG (encoding filaggrin) are a major predisposing factor for atopic dermatitis. A number of association studies in populations of diverse ancestry, meta-analyses of such studies and GWASs have reported that mutations in FLG are strongly associated with atopic dermatitis. Apart from FLG, recent GWASs of European and Chinese populations and a meta-analysis of GWASs have identified six susceptibility loci of atopic dermatitis with genome-wide significance.

The first GWAS of atopic dermatitis identified a
susceptible locus on chromosome 11q13.5, located 38 kb downstream of C11orf30.46 The locus contained LRRC32. LRRC32 mRNA is highly expressed in activated FOXP3+ regulatory T cells, and LRRC32 is essential for the surface expression of latent TGF-β on the cells.49 In the GWAS, an additional association signal was observed in the epidermal differentiation complex on chromosome 1q21. The region is located adjacent to FLG and contains a number of genes that play roles in epidermal structure and function.50

A GWAS for atopic dermatitis in the Chinese Han population identified two new susceptibility loci at 5q22.1 and 20q13.33 and confirmed an association between the FLG locus and atopic dermatitis at the genome-wide level of significance.47 The 5q22.1 locus contains TMEM232 and SLC25A46, but little is known about the functions of the genes.47 Interestingly, the region is located adjacent to the TSLP gene; however, the influence of the associated variants in the region on the expression of the TSLP gene remains unclear. The 20q13.33 region contains eight genes, including the TNFRSF6B gene, which encodes decoy receptor 3 (DcR3).47 DcR3 is a secreted protein belonging to the tumor necrosis factor receptor family (TNFRSF) and it binds to LIGHT. A recent study has shown that LIGHT is a target for asthmatic airway remodeling.51

In 2011, a meta-analysis of GWAS reported three novel susceptibility loci for atopic dermatitis.48 The peak associations were observed at rs479844 upstream of OVOL1 and rs2164983 near ACTL9, both of which are near genes that have been implicated in epidermal proliferation and differentiation, as well as rs2897442 in KIF3A at 5q31.1.48 The locus at 5q31 contains a cluster of cytokine and immune-related genes and has been associated with several autoimmune and inflammatory diseases including asthma and psoriasis.24,52 Interestingly, conditional logistic regression analysis indicated that the association signal within the cytokine cluster at 5q31.1 appeared to be composed of two independent signals, one centered at IL13-RAD50 and the other at IL4-KIF3A.48 These findings suggest that the genetic region containing IL4 and IL13 influences the development of both atopic dermatitis and asthma.

In 2012, we reported a GWAS and a validation study comprising a total of 3,328 cases and 14,992 controls in the Japanese population.53 Analyzing a total of roughly 600,000 genetic variants, a total of eight new genetic regions associated with atopic dermatitis were identified.53 The study confirmed the seven loci observed in earlier studies.53 The associated region at 2q12 contains the receptors of IL-1 family cytokines IL1RL1, IL18R1 and IL18RAP. IL-1 family members are abundantly expressed in the skin, and IL1RL1 is a component of the IL-33 receptor, which is expressed by Th2 cells and mast cells.25,54 Interestingly, IL-33 is secreted in damaged tissues of atopic dermatitis and promotes Th2-type immune responses and the pathogenesis of atopic dermatitis.25 The most significant association at the locus is found in the MHC class III region. In the GWAS, after conditional analysis, there were two independent association signals in the MHC class I and class III regions. The involvement of autoimmunity has been suggested in chronic inflammation in patients with atopic dermatitis.42 IgE antibodies against keratinocytes and endothelial cells are observed in serum specimens from subjects with severe atopic dermatitis.55 Since the MHC is associated with a number of autoimmune diseases,56 further experiments will be required to clarify where the susceptibility genes are located precisely. The region of association at 11p15.4 contains NLRP10, which belongs to the NALP protein family but lacks the leucine-rich repeat region. A recent report has shown that NLRP10 is essential to initiate adaptive immunity by dendritic cells.57 Another study has shown that NLRP10 plays a role in the control of disseminated C. albicans infection in vivo through the generation of adaptive immune responses such as Th1 and Th17 responses, to fungal infection.58 The chromosome 3p21.33 region is located adjacent to the CCR4 gene, which encodes a Th2-associated chemokine receptor for CCL22 and CCL17 (TARC). CCR4 mediates skin-specific recruitment of T cells during inflammation.42,59 Keratinocyte-derived TSLP induces dendritic cells to produce TARC, and serum TARC levels are useful for evaluation of the disease activity of AD.42,59 The associated region at 3q13.2 contains CCDC80, and CCDC80 is involved in induction of C/EBPα and peroxisome proliferator-activated receptor (PPAR) γ.60 C/EBP α is expressed in basal keratinocytes and upregulated while keratinocytes exit the basal layer and undergo terminal differentiation.61 The associated region at 7p22 contains CARD11, which encodes CARMA1, an essential scaffold protein for lymphocyte activation via T cell receptor and B cell receptor signaling.62 CARMA1 has a critical role in the regulation of JunB and GATA3 transcription factors and subsequent production of Th2 cell-specific cytokines.63 Interestingly, mice homozygous for the Carma-1 mutation gradually develop atopic dermatitis with hyper-IgE.64 The region of association at 10q21.2 contains EGR2, a T cell anergy-associated transcription factor that activates the expression of genes involved in the negative regulation of T cell proliferation and inflammation.55 The associated region at 20q13 includes CYP24A1, which encodes a mitochondrial cytochrome p450 superfamily enzyme. The protein acts as a degradation enzyme of 1,25-dihydroxyvitamin D3, the active form of vitamin D3.66 Vitamin D is a modulator of innate and adaptive immune system functions, and plays a crucial role in production of antimicrobial peptides in skin.66 A recent study has shown an association between vitamin D deficiency and the severity of atopic dermatitis.67 Topical glucocorticoids
The allergic march is the natural history of allergies. Symptoms often appear in a particular sequence during childhood. Figure 1 shows the sequence of symptoms: about age of 0-2 years, food allergy, atopic dermatitis; age of 7 years, bronchial asthma, allergic rhinitis; age of 12 years, inhalation allergies, skin sensitization, and specific food allergies. Genetic factors include the regions 2q12 (IL1RL1/IL18R1/IL18RAP), 6p21.3 (MHC region), 5q31 (KIF3A/IL4/IL13), 2q12 (IL1RL1/IL18R1/IL18RAP), 6p21.3 (MHC region), and 11q13.5 (C11orf30/LRRC32). The 11q13.5 region is associated with both allergic rhinitis and atopic dermatitis.

**GWASs OF ALLERGIC RHINITIS AND EOSINOPHILIC ESOPHAGITIS**

A large-scale genome-wide meta-analysis identified several loci associated with allergic rhinitis and grass sensitization.68 The chromosome 11q13.5 region, which is near C11orf30 and LRRC32, is associated with both phenotypes with genome-wide significance \((P < 5 \times 10^{-8})\).68 LRRC32 is a key receptor controlling FOXP3 in human regulatory T cells, which are critical for maintaining tolerance.69 The 11q13.5 region has already been reported as a susceptibility locus for atopic dermatitis and asthma.38,46 In the GWAS, the HLA region was also associated with grass sensitization at the genome-wide level of significance.68

Eosinophilic esophagitis (EoE) is inflammation of the esophagus with abnormal infiltration of eosinophils in an allergic reaction. A recent GWAS reported that a candidate locus at chromosome 5q22 contained TSLP and was associated with EoE.70 Interestingly, recent GWASs have identified an association of the TSLP region with bronchial asthma.27,29 In one study, TSLP was found to be overexpressed in esophageal biopsies from individuals with eosinophilic esophagitis compared with unaffected individuals, and TSLP expression levels were significantly correlated with the disease-related rs3806932 SNP in EoE cases.70

**GENETIC COMPONENTS OF ALLERGIC MARCH**

Recent GWASs have revealed shared immunological mechanisms in several immune-related diseases.71 A number of GWASs of allergic diseases have revealed that different allergic diseases share overlapping susceptibility loci. Among regions identified by GWASs with genome-wide significance, the IL-1 receptor cluster region on 2q12, the HLA and IL13 regions were found to be associated with both atopic dermatitis and bronchial asthma.24,29,48,53 The C11orf30/LRRC32 region on chromosome 11q13.5 is associated with atopic dermatitis, bronchial asthma, and allergic rhinitis.38,46,68 The allergic march is the natural history of allergies and their symptoms often appearing in a particular sequence during childhood, and atopic dermatitis often represents the beginning of this allergic march (Fig. 1).72 Recent studies have shown that about half of the patients with atopic dermatitis will develop bronchial asthma, particularly with severe atopic dermatitis, and two-thirds will develop allergic rhinitis.72 Since patients with the allergic march often show a severe phenotype, further studies of these overlapping loci (Fig. 1) might help to clarify the mechanisms underlying this phenomenon.

**FUTURE PERSPECTIVES**

Although GWASs have identified disease susceptibility loci of human diseases and provided valuable insights into their genetic components, they have explained little of heritability, and the associated variants have small effect sizes.73 Potential sources of missing heritability include unmapped common and rare variants, copy number variations, epigenetic ef-
fects, gene-gene interactions and gene-environment interactions. Recently, the Immunochip has been developed to conduct deep replication of major autoimmune and inflammatory diseases, and fine-mapping of established loci identified by GWASs. The Immunochip is a custom Illumina Infinium high-density array containing 196,524 polymorphisms selected using data from the 1000 Genomes Project and other available disease-specific resequencing data. The Immunochip approach might improve our understanding of allergic diseases; however, the Immunochip does not cover the whole genome and is designed for use in white European populations. It would seem to be less informative for Asian populations.

The 1000 Genomes Project is the first project to sequence the genomes of a large number of people to provide a comprehensive resource on human genetic variation. The project aims to find essentially all variants with frequencies >1% across the genome and >0.1% in protein-coding regions. It is considered that imputations based on the 1000 Genomes Project can increase the opportunity to find novel association signals in GWAS meta-analyses through the dense marker map and large number of haplotypes. Since various GWAS consortiums for allergic diseases have recently been formed, meta-analyses of GWASs using the 1000 Genome Project data will identify novel disease-associated regions.

Candidate genes at the susceptible loci identified by GWASs of allergic diseases suggest roles for epidermal barrier functions, innate and adaptive immunity, IL-1 signaling, the inflammatory response, regulatory T cells and the vitamin D pathway in the pathogenesis of allergic disorders. In immunology, consortium biology has been conducted to establish complete charts of all the molecular components that play roles in cells of the immune system. To construct animal models that mimic human physiology, the findings of GWASs will be helpful to highlight the genes involved in human allergic diseases.

Epigenetic regulation plays an important role in mediating environmental influences on gene expression. Promoter methylation, histone tail modifications and altered expression of non-coding RNAs influence gene regulation and are candidate targets for genetic associations. A recent study reported sequence-dependent allele-specific DNA methylation and that cis-regulatory variants influence gene expression and affect chromatin states. Sequence variants can influence communication between different parts of the genome, and SNPs might alter chromatin networks in a genotype-specific manner. It is known that the human genome encodes 100 evolutionarily conserved families of micro RNAs (miRNAs), and SNPs can also directly affect miRNA binding sites. Further epigenetic analysis using GWAS SNPs and their LD regions are required.

CONCLUSION

In the past century, most biological research has been hypothesis-driven investigation performed in individual laboratories. Recent GWASs and meta-analyses of the that comprehensively assess genes related to multifactorial diseases in a non-biased manner across the whole genome have enhanced our understanding of human allergic diseases. Further cross-disciplinary studies combining genetics, immunology, epidemiology, and clinical allergology are necessary for translation of research into clinical practice. It is anticipated that the cross-disciplinary studies will help to protect humans from developing allergic diseases and provide molecular targets for therapeutic intervention.

REFERENCES

17. Hirota T, Harada M, Sakashita M et al. Genetic polymorphism regulating ORM1-like 3 (Sacharomycyes cerevisiae) expression is associated with childhood atopic


